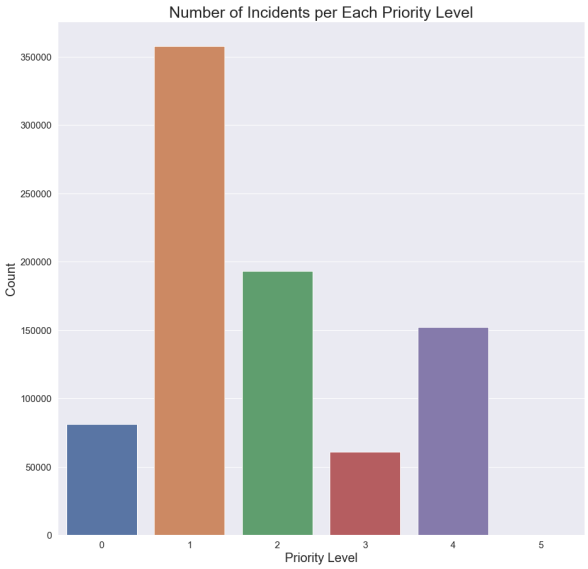
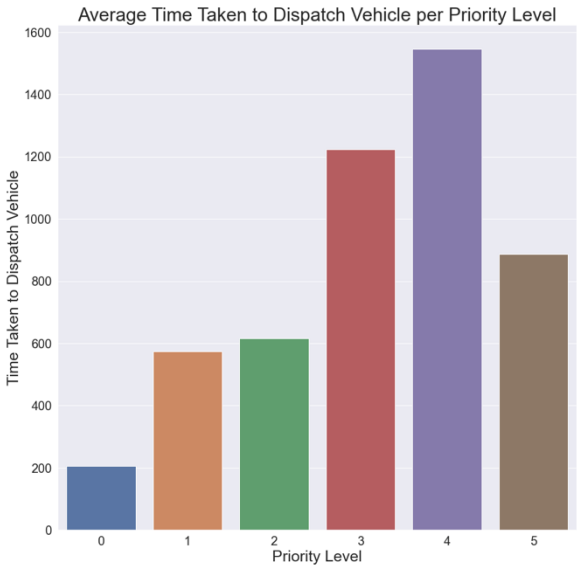
Rohin Garg
Professor Iapalucci
DATA 205
14 May 2021

# Analysis of Collisions and Police Incidents

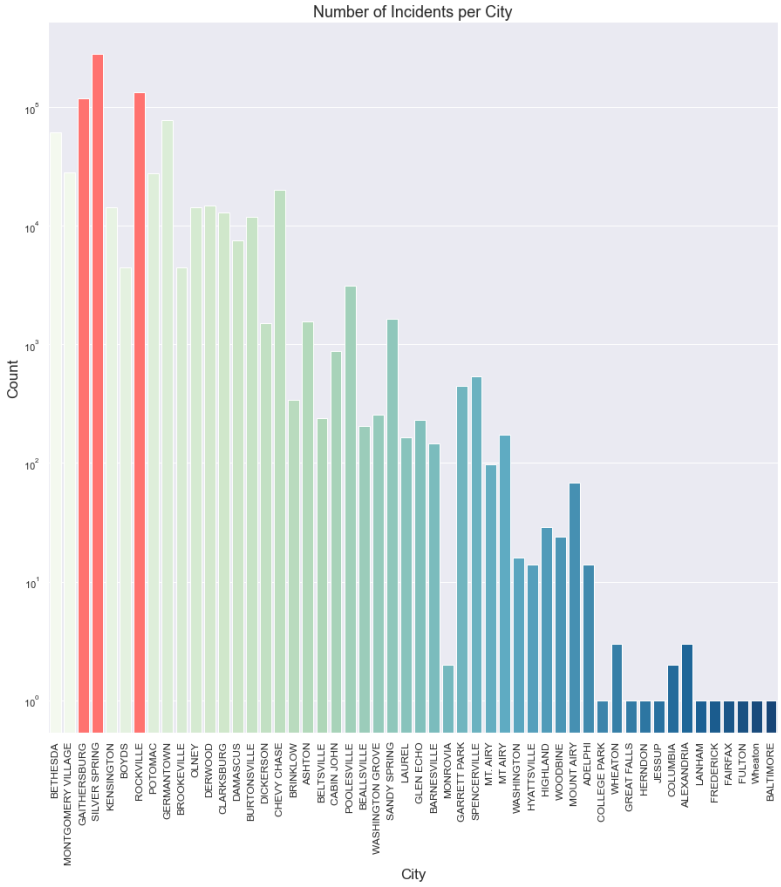
To conduct my analysis, I used three datasets: *Police Dispatched Incidents, Crash Reporting – Incidents Data,* and *Census Data*. The *Police Dispatched Incidents* dataset was imported from dataMontgomery and contains a list of records showing every incident where police had to be dispatched dating back to April 2nd, 2017. This dataset is updated four times daily and shows 805 thousand incidents. One variable of interest is "Priority". This variable is a variable from 1-5 which indicates the priority level of the incident. As the priority level gets smaller, the priority of the incident is higher. Therefore, a priority level of 0 is of the utmost priority whereas a priority level of 5 is the lowest priority. Another variable of interest is "Calltime Dispatch", a continuous variable showing the time taken for a police officer to be dispatched to the incident. The second dataset, *Crash Reporting – Incidents Data,* was also imported from dataMontgomery and provides general information about each collision and details of all traffic collisions occurring on county and local roadways within Montgomery County dating back to January 1st, 2015. The dataset is updated weekly and contains 69.9 thousand collisions. One variable of interest is "Hit/Run". For each collision, this variable is either yes or no, with yes indicating a hit-and-run collision. Another variable of interest is "Collision Type", which is also a categorical variable and classifies the collision. The third dataset used was imported from the Census Bureau. The dataset was imported to show the total population and population of different age ranges for each ZIP Code in Maryland.

When initially looking at the *Police Dispatched Incidents* data, I wondered how often are incidents classified as a certain priority level and how does the priority level affect dispatch time?



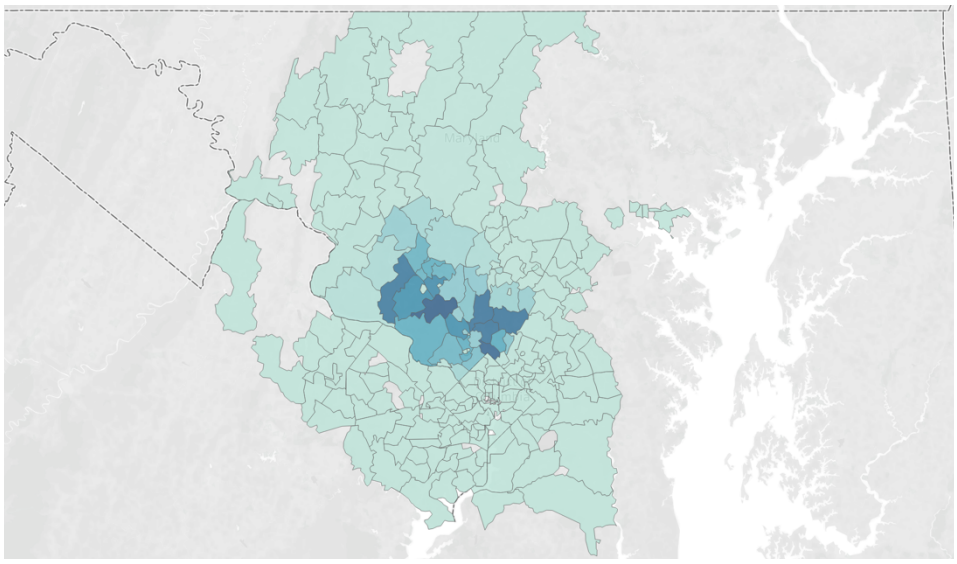



The visualization on the left shows that most incidents were classified as a priority level 1. Also, there are a very little number of incidents where police were dispatched that were classified as a level 5. The visualization on the rights shows a trend where the dispatch time decreased as the priority level decreased. This is exactly what we would want to see since it means that higher the importance of an incident, the quicker a police officer was dispatched. However, the dispatch time for priority level 5 shows a discrepancy from the trend. We can attribute this discrepancy to the fact that there weren't enough incidents with the priority level 5 which skewed the data.

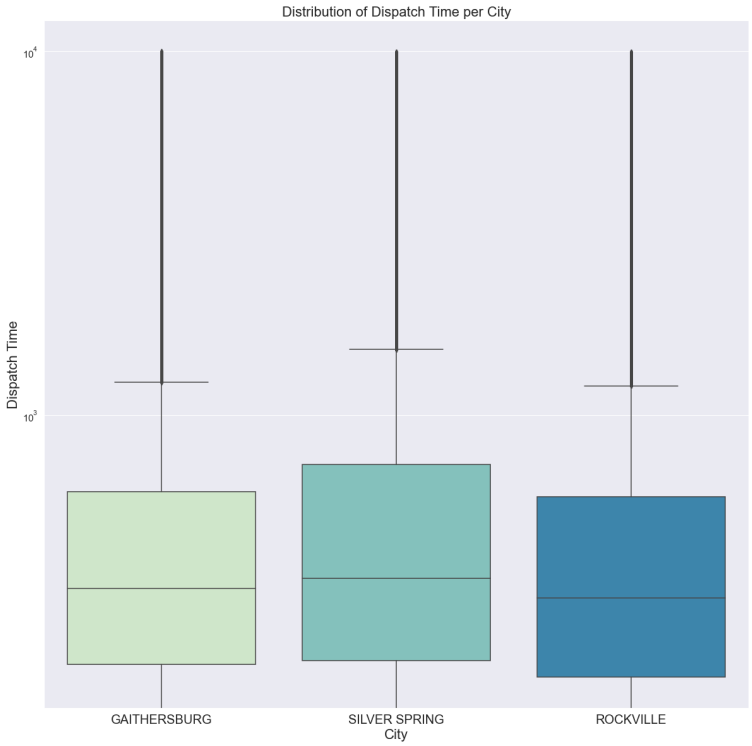



The next question I analyzed was which cities have the most incidents. The three cities with the most incidents are Silver Spring, Rockville, and Gaithersburg. They are highlighted in red in the visualization to the right. Keep in mind that Y-Axis is logarithmically scaled, so the difference in incidents between cities is actually larger than pictured.

The map below shows the number of incidents per ZIP Code. The darker the region, the more incidents have occurred.



When looking at this dataset, I wanted to figure out if police officers tend to be dispatched earlier or later to certain cities. To do this analysis, I looked at the mean dispatch times for the three cities which had the most incidents: Silver Spring, Rockville, and Gaithersburg. To start, I created a boxplot for each city to compare the distributions for the dispatch time.

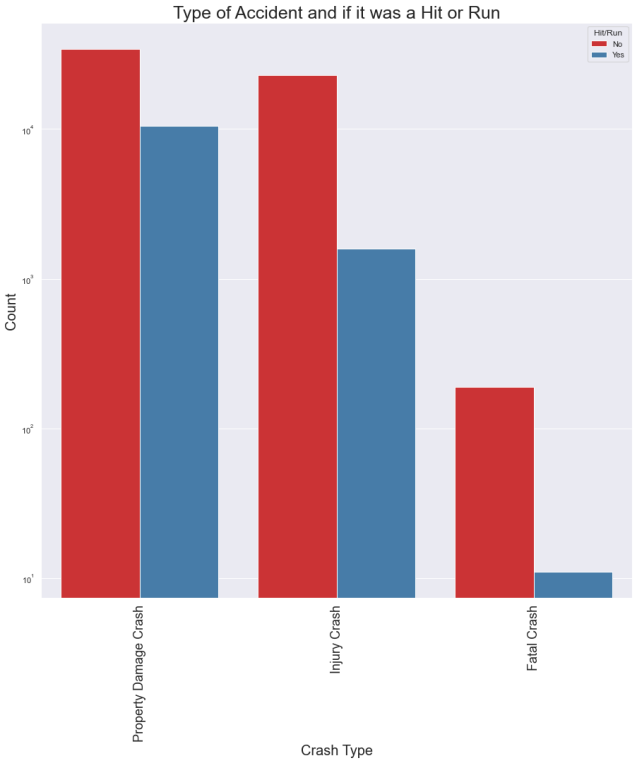



As shown in the boxplot, Silver Spring has the highest average dispatch time, with Gaithersburg as second and Rockville as third. To determine if the difference between the average dispatch times were statistically significant, I conducted a Chi-squares test which resulted in a p-value less than 0.05. Therefore, we can reject the null hypothesis which states that the average dispatch times are equal. We safely conclude that if you live in Silver Spring, you will tend to experience longer dispatch times than if you live in Rockville or Gaithersburg.

Next, I looked at how often a collision resulted in a hit-and-run depending on three types of crashes: a property damage crash, injury crash, and fatal crash.

NOTE: The Y-Axis is logarithmically scaled. The red bar indicates that there was no hit-and-run whereas the blue bar indicates there was.

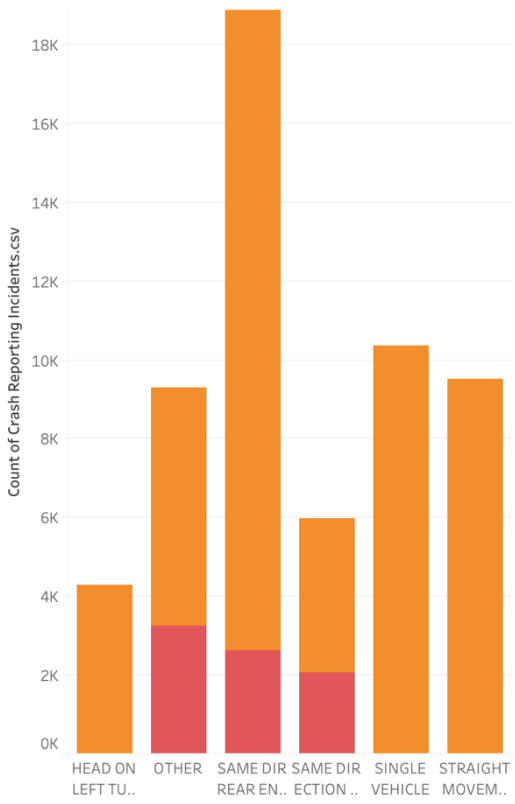
The probability of a hit-and-run was relatively high when the crashes resulted in property damage compared to when the crashes resulted in injury or death.



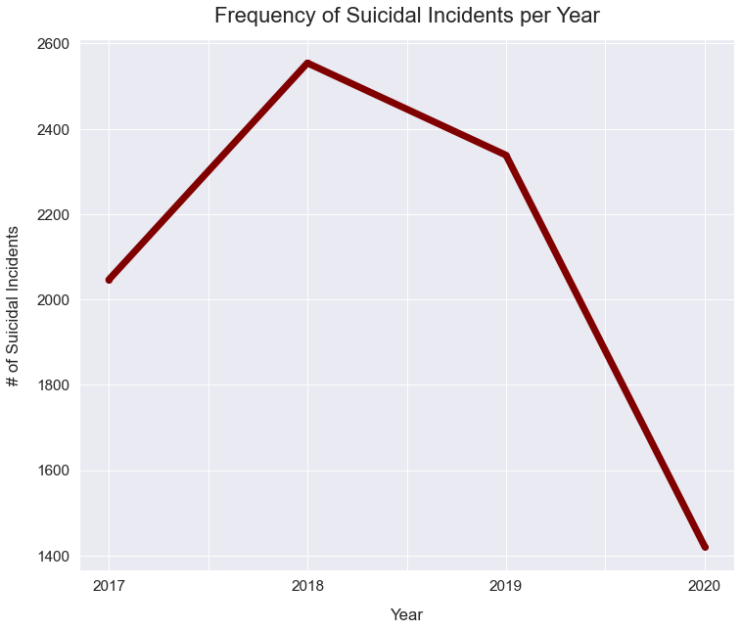
To further analyze which collisions caused a hit-and-run, I created a stacked bar chart.

NOTE: The red part represents the collisions which resulted in a hit-and-run.

The visualization shows that the type of collisions which most resulted in a hit-and-run was a “same direction rear end”. This means that the likelihood of a hit-and-run occurring was highest in rear end collisions when travelling in the same direction

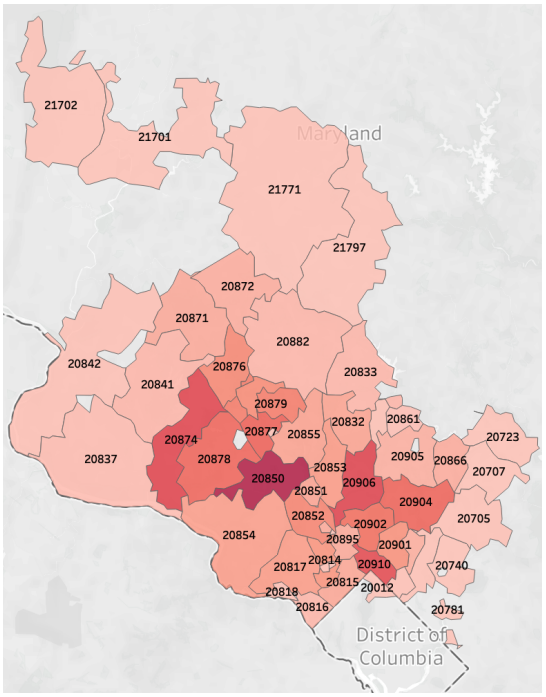


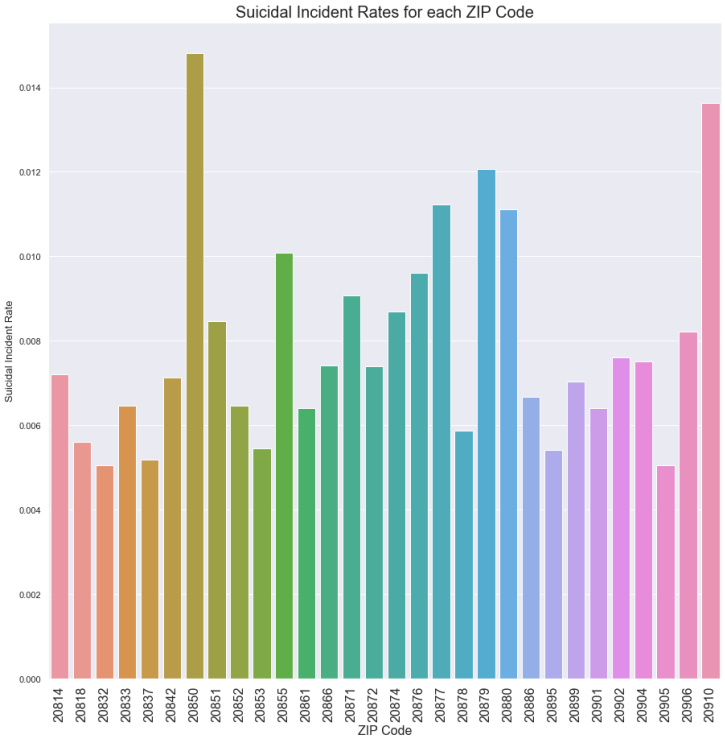

The next part of my analysis focuses on suicidal incidents in Montgomery County from 2017 to 2020. First, I looked at how the frequency of reported suicidal incidents has changed over the past four years.




Surprisingly, the frequency of suicidal incidents dropped significantly in 2020. Many had speculated that the quarantine would lead to a severe lack of social interaction, causing depression, and then ultimately resulting in an increase in suicidal incidents. However, that was not the case. The number of suicidal incidents in 2020 dropped nearly 50 percent from 2018.

To continue the analysis of suicidal incidents, I wanted to figure out where the highest frequency of suicidal incidents took place. To do that, I looked at which ZIP codes had the highest total number of suicidal incidents and which ZIP codes had the highest rates of suicidal incidents based off their population.

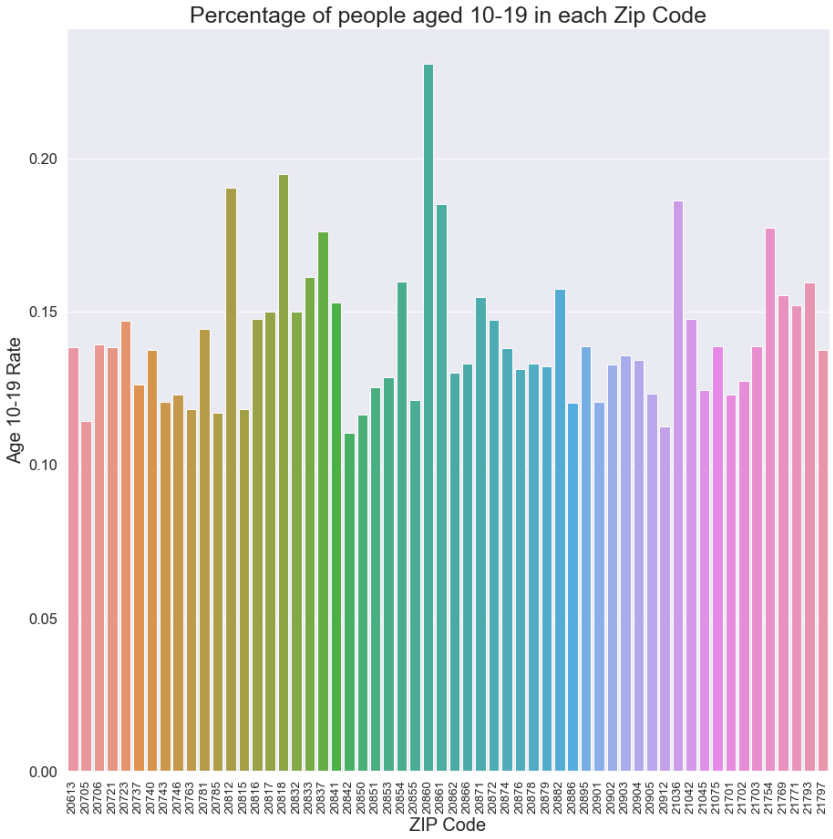


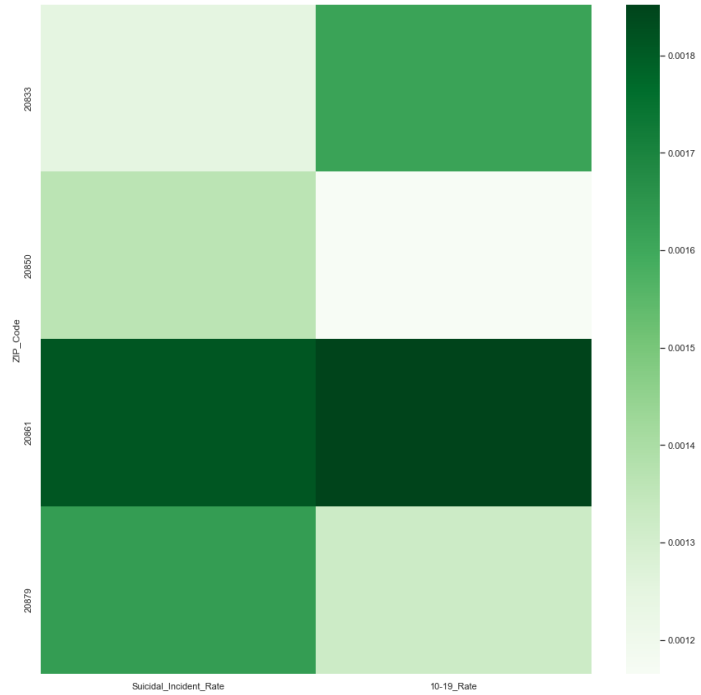

The map on the left shows the frequency of total suicidal incidents where police were dispatched in the past four years. The darker regions represent a higher number of suicidal incidents whereas the lighter regions represent less. Although this data was taken from the Montgomery County Police, there are some ZIP codes outside Montgomery County that are shaded because Montgomery County police officers were dispatched there.

Rather than just determining the total frequency of suicidal incidents in each ZIP code, I used the data provided by the Census to calculate the suicidal incident rate for each ZIP Code, calculated by (# of suicidal incidents / population of ZIP code). As shown in the bar chart on the right, the ZIP code with the highest suicidal incident rate is 20850 which is in the Rockville area. The second highest is 20910 which is in the Silver Spring and Takoma Park area.

In relation to the suicidal incident rates, I examined the percentage of teenagers in the total population for each ZIP code to see if that was correlated to the suicidal incident rate. The percentage of teenagers in the population was calculated by (# of citizens aged 10-19 / total population) for each ZIP Code. The data was imported from the Census Bureau. The bar chart below shows the rate of teenagers for each ZIP code.




To visualize the correlation between the rate of teenagers and the suicidal incident rate for ZIP codes, I created a heatmap. The left side represents the suicidal incident rate, and the right side represents the rate of teenagers. The darker hue of green represents a higher rate of both teenagers and suicidal incident rates. The heatmap shows that a higher suicidal incident rate correlates with a higher rate of teenagers, and that a lower suicidal incident rate correlates with a lower rate of teenagers. Although we cannot say for certain that teenagers cause a higher suicidal incident rate, there is a trend shown that it does in Montgomery County in the past four years. A policy that could be implemented to try to prevent this trend would be to establish more mental health and suicide prevention programs throughout the K-12 education system.



Although both datasets that I used throughout this project were of high quality, I have some recommendations the enhance the datasets to allow data scientists to have a deeper analysis with this data. The two datasets I used from dataMontgomery was the *Police Dispatched Incidents* dataset and the *Crash Reporting – Incidents Data* dataset. I would recommend for the *Police Dispatched Incidents* dataset to add variables regarding the victim of the incident. Therefore, one could create more inferences from the incidents. In my case, I would've used the ages of the victims in the suicidal incidents to investigate deeper on whether children specifically have a higher rate of suicidal incidents. I would also recommend adding a note explaining that a lower Priority Level means that the actual priority is higher. I was confused about this and had it backwards for the first couple weeks working on the project. For the *Crash Reporting Incidents* dataset, I would recommend adding a column depicting the type of vehicle or vehicles involved in the collision. Therefore, one can analyze if certain car models or car brands have a higher rate of collision over a specific time period.

**Thank you so much to…**

- The Montgomery County Government
- Victoria Lewis
- dataMontgomery
- Professor Iapalucci
- Professor Saidi
- Professor Abdirisak
- Fellow Students

**for your support and direction throughout this project!**